Infection, Genetics and Evolution 32 (2015) 416–424

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Genomic and single nucleotide polymorphism analysis of infectious bronchitis coronavirus

Celia Abolnik *

Poultry Section, Department of Production Animal Studies, Faculty of Veterinary Science, University of Pretoria, Onderstepoort 0110, South Africa

ARTICLE INFO

Article history:
Received 5 December 2014
Received in revised form 13 March 2015
Accepted 26 March 2015
Available online 3 April 2015

Keywords:
Infectious bronchitis
Coronavirus
SNP
Spike
HVR

ABSTRACT

Infectious bronchitis virus (IBV) is a *Gammacoronavirus* that causes a highly contagious respiratory disease in chickens. A QX-like strain was analysed by high-throughput Illumina sequencing and genetic variation across the entire viral genome was explored at the sub-consensus level by single nucleotide polymorphism (SNP) analysis. Thirteen open reading frames (ORFs) in the order 5′-UTR-1a-1ab-S-3a-3b-E-M-4b-4c-5a-5b-N-6b-3′UTR were predicted. The relative frequencies of missense: silent SNPs were calculated to obtain a comparative measure of variability in specific genes. The most variable ORFs in descending order were E, 3b, 5′UTR, N, 1a, S, 1ab, M, 4c, 5a, 6b. The E and 3b protein products play key roles in coronavirus virulence, and RNA folding demonstrated that the mutations in the 5′UTR did not alter the predicted secondary structure. The frequency of SNPs in the Spike (S) protein ORF of 0.67% was below the genomic average of 0.76%. Only three SNPS were identified in the S1 subunit, none of which were located in hypervariable region (HVR) 1 or HVR2. The S2 subunit was considerably more variable containing 87% of the polymorphisms detected across the entire S protein. The S2 subunit also contained a previously unreported multi-A insertion site and a stretch of four consecutive mutated amino acids, which mapped to the stalk region of the spike protein. Template-based protein structure modelling produced the first theoretical model of the IBV spike monomer. Given the lack of diversity observed at the sub-consensus level, the tenet that the HVRs in the S1 subunit are very tolerant of amino acid changes produced by genetic drift is questioned.

1. Introduction

Coronaviruses (family *Coronaviridae*, order *Nidovirales*) are enveloped, single-stranded RNA viruses with large genome sizes of ~25–30 kb. The family is split into four genera: *Alpha-*, *Beta*, *Gamma* and *Deltacoronaviruses*, each containing pathogens of veterinary or human importance. A current evolutionary model postulates that bats are the ancestral source of *Alpha-* and *Betacoronaviruses* and birds the source of *Gamma-* and *Deltacoronaviruses* (Woo et al., 2012). The *Alphacoronaviruses* infect swine, cats, dogs and humans. *Betacoronaviruses* infect diverse mammalian species including bats, humans, rodents and ungulates. The SARS coronavirus (SARS-CoV), which verged on a pandemic in 2003 with 8273 cases in humans and 755 deaths is a *Betacoronavirus*. Another member of this genus, the recently-discovered Middle East Respiratory Syndrome (MERS) coronavirus (MERS-CoV) has claimed 88 human lives from 212 cases since April 2012, and dromedary camels are the suspected reservoir (Briese et al., 2014). Genus *Gammacoronavirinae* includes strains infecting birds and whales (Woo et al., 2012; McBride et al., 2014; Borucki et al., 2013) and deltacoronaviruses have been described in birds, swine and cats (Woo et al., 2012). The diversity of hosts and genomic features amongst CoVs have been attributed to their unique mechanism of viral recombination, a high frequency of recombination, and an inherently high mutation rate (Lai and Cavanagh, 1997).

Infectious bronchitis virus (IBV) is a gammacoronavirus which causes a highly contagious respiratory disease of economic importance in chickens (Cook et al., 2012). IBV primarily replicates in the respiratory tract but also, depending on the strain, in epithelial cells of the gut, kidney and oviduct. Clinical signs of respiratory distress, interstitial nephritis and reduced egg production are common, and the disease has a global distribution (Cavanagh, 2007; Cook et al., 2012). The IBV genome encodes at least ten open reading frames (ORFs) organised as follows: 5′ UTR-1a-1ab-S-3a-3b-E-M-5a-5b-N-3a-3′UTR. Six mRNAs (mRNA 1–6) are associated with production of progeny virus. Four structural proteins including the spike glycoprotein (S), small membrane protein (E), membrane

* Address: University of Pretoria, Private Bag X04, Onderstepoort 0110, South Africa. Tel.: +27 12 529 8258.
E-mail address: celia.abolnik@up.ac.za

http://dx.doi.org/10.1016/j.meegid.2015.03.033
1567-1348/

glycoprotein (M), and nucleocapsid protein (N) are encoded by mRNAs 2, 3, 4 and 6, respectively (Casais et al., 2005; Hodgson et al., 2006). Messenger RNA (mRNA) 1 consists of ORF1a and ORF1b, encoding two large polyproteins via a ribosomal frameshift mechanism (Inglis et al., 1990). During or after synthesis, these polyproteins are cleaved into 15 non-structural proteins (nsp2–16) which are associated with RNA replication and transcription. The S glycoprotein is post-translationally cleaved at a protease cleavage recognition motif into the amino-terminal S1 subunit (92 kDa) and the carboxyl-terminal S2 subunit (84 kDa) by the host serine protease furin (de Haan et al., 2004). The multimeric S glycoprotein extends from the viral membrane, and the globular S1 subunit is anchored to the viral membrane by the S2 subunit via non-covalent bonds. Proteins 3a and 3b, and 5a and 5b are encoded by mRNA 3 and mRNA 5, respectively and are not essential to viral replication (Casais et al., 2005; Hodgson et al., 2006).

A confounding feature of IBV infection is the lack of correlation between antibodies and protection, and discrepancies between *in vitro* strain differentiation by virus neutralization (VN) tests and *in vivo* cross-protection results. Taken with the ability for high viral shedding in the presence of high titres of circulating antibodies, the involvement of other immune mechanisms are evident, and the roles of cell-mediated immunity and interferon have been experimentally demonstrated (Timms et al., 1980; Collisson et al., 2000; Pei et al., 2001; Cook et al., 2012).

Dozens of IBV serotypes that are poorly cross-protective have been discovered and studied by VN tests and molecular characterisation of the S protein gene. Most of these serotypes differ from each other by 20–25% at amino acid level in S1, but may differ by up to 50%. S1 contains the epitopes involved in the induction of neutralizing, serotype-specific and hemagglutinaton inhibiting antibodies (Cavanagh, 2007; Darbyshire et al., 1979; Farsang et al., 2002; Ignjatovic and McWaters, 1991; Meulemans et al., 2001; Gelb et al., 1997). Most of the strain differences in S1 occur in three hypervariable regions (HVRs) located between the amino acid residues 56–69 (HVR1), 117–131 (HVR2) and 274–387 (HVR3) (Moore et al., 1997; Wang and Huang, 2000). Monoclonal antibody analysis mapped the locations of many of the amino acids involved in the formation of VN epitopes to within the first and third quarters of the linear S1 polypeptide (De Wit, 2000; Kant et al., 1992; Koch et al., 1990), which is where closely-related stains (>95% amino acid identity) also differ (Bijlenga et al., 2004; Farsang et al., 2002). Cavanagh (2007) proposed that these parts of the S1 subunit are very tolerant of amino acid changes, conferring a selective advantage. Recently, the receptor-binding domain of the IBV M41 strain was mapped to residues 16–69 of the N terminus of S1, which overlaps with HVR1 (Promkuntod et al., 2014).

The S2 subunit, which drives virus-cell fusion, is more conserved between serotypes than S1, varying by only 10–15% at the amino acid level (Bosch et al., 2005; Cavanagh, 2005). Although it was initially thought that S2 played little or no role in the induction of a host immune response, it has since been shown that an immunodominant region located in the N-terminal half of the S2 subunit can induce neutralizing, but not serotype-specific, antibodies demonstrated by the ability of this subunit to confer broad protection against challenge with an unrelated serotype (Kusters et al., 1989; Toro et al., 2014).

IBVs are continuously evolving as a result of (a) frequent point mutations and (b) genomic recombination events (Cavanagh et al., 1992; Kottier et al., 1995; Jackwood et al., 2005; Zhao et al., 2013; Kuo et al., 2013; Liu et al., 2014). Multiple studies on IBV diversity have focused on inter-serotypic and inter-strain variation, and a few have focused on sub-populations within the S1 subunit in vaccine strains (Gallardo et al., 2012; Ndegwa et al., 2014). The present study aimed to explore genetic variation across the entire viral genome at the sub-consensus level. It was anticipated, based on the published literature, that certain regions, and the S1 subunit HVRs in particular, would display significant sub-genomic variation. This study focused on a QX-like strain, a serotype currently causing significant poultry health problems across Europe, Asia, South America and South Africa.

2. Materials and methods

2.1. Origin and isolation of QX-like strain ck/ZA/3665/11

Twenty-eight-day old chickens in a commercial broiler operation presented with acute lethargy, reduced feed consumption and mortality. Tracheitis and swollen kidneys were noted on post mortem, as well as a secondary *Escherichia coli* infection. The worst affected houses had mortality rates of 19.8%, 11.9% and 10.2%. IBV was isolated in specific pathogen free (SPF) embryonated chicken eggs (ECE) as described in Knoetze et al. (2014). After an initial two passages in ECE, the virus was passaged twice further at the University of Pretoria.

2.2. Preparation of the genome and Illumina sequencing

RNA was extracted from allantoic fluid using TRIzol® reagent (Ambion, Life Technologies, Carlsbad, USA) according to the manufacturer's protocol. The genome was transcribed to cDNA and amplified using a TransPlex® Whole Transcriptome Amplification kit (Sigma–Aldrich, Steinheim, Germany). Illumina MiSeq sequencing on the cDNA library was performed at the ARC-Biotechnology Platform, Onderstepoort, Pretoria.

2.3. Genome assembly, RNA folding and recombination analysis

Illumina results were analysed using the CLC Genomics Workbench v 5.1.5. Paired-end reads were trimmed and a preliminary *de novo* assembly was performed. The larger segments were analysed by BLAST to identify the closest genomic reference strain (ITA/90254/2005, CAZ86699). This strain was retrieved and used as a scaffold for assembly-to-reference, generating a consensus sequence for 3665/11. Trimmed paired-end reads were also mapped against other IBV serotype genomes, subsequently confirming that strain 3665/11 was a pure culture of a QX-like IBV. The genome was deposited in Genbank under the accession number KP662631. RNA folding was predicted using the CLC Genomics Workbench v 5.1.5. Genetic recombination in the consensus sequence was evaluated using the recombination detection program RDP v4.31.

2.4. Genome annotation and single nucleotide polymorphism (SNP) analysis

Coding sequence and ORF prediction was carried out in VIGOR (Wang et al., 2010). Trimmed paired-end reads were re-mapped against the 3665/11 consensus sequence for SNP detection. A SNP detection table generated in the CLC Genomics Workbench was manually edited to eliminate all SNPs with a frequency of <5%. This conservative cutoff was selected to eliminate any non-specific PCR errors introduced during preparation of the transcriptome library or deep sequencing, and excluded most of the point insertions producing gaps and frameshift mutations across the genome. Nucleotide substitutions in coding regions were manually inspected for changes to the consensus amino acid (Table 1, Supplementary data). Motifs were predicted using the ELM Eukaryotic Linear Motif Resource for Functional Sites in Proteins (Dinkel et al., 2014).